# Uncovering randomness and success in society

**Sarika Jalan**[1,2,*], **Camellia Sarkar**[2], **Anagha Madhusudanan**[1], **Sanjiv Kumar Dwivedi**[1]

[1] *Complex Systems Lab, Physics Discipline, Indian Institute of Technology Indore, M-Block, IET-DAVV Campus, Khandwa Road, Indore 452017, India*

[2] *Complex Systems Lab, Center for Biosciences and Biomedical Engineering, Indian Institute of Technology Indore, M-Block, IET-DAVV Campus, Khandwa Road, Indore 452017, India*

[*] *E-mail: sarika@iiti.ac.in*



**Abstract**

An understanding of how individuals shape and impact the evolution of society is vastly limited due to the unavailability of large-scale reliable datasets, that can simultaneously capture information regarding individual movements as well as social interactions. We believe that the popular Indian film industry, 'Bollywood', can provide a social network apt for such a study. Bollywood provides massive amounts of real, unbiased data that spans over 100 years and hence this network has been used as a model for the present paper. It is seen that the nodes which maintain a moderate degree or widely cooperate with the other nodes of the network, tend to be more fit (measured as the success of the node in the industry) in comparison to the other nodes. The analysis carried forth in the current work, using a conjoined framework of complex network theory and random matrix theory, aims to quantify the elements that determine the fitness of an individual node and the factors that contribute to the robustness of a network. The authors of this paper believe that the method of study used in the current paper can be extended to study various other industries and organizations.

## 1 Introduction

The field of network analysis helps us to look at the study of an individual component as a part of a complex social structure and its interactions [8]. It explains various phenomena in a wide variety of disciplines ranging from physics to psychology to economics. The theory is adept at finding the causal relationships between network attributes such as the position of a node and the specific ties associated with it, and the fitness of the said node [2]. Such relationships, that seemed thoroughly random to the eyes of a researcher only about a decade before, have now been vastly studied and documented [3]. We aim to further investigate the very interesting idea that human behavior is predictable to a fair degree [4] using the Bollywood Network as a model for this purpose.

Making nearly one thousand feature films and fifteen hundred short films per year, the Indian film industry is the largest in the world [5] which has held a large global population in more spheres of its existence than just entertainment. It mirrors a changing society capturing its peaks and valleys over time and impacts the opinions and views of the diverse populace [6]. An example that can be stated as a proof of this was exhibited when the number of Indian tourists to Spain increased by $65\%$ in the year succeeding the box office success of the movie 'Zindagi Na Milegi Dobara', which extensively portrayed tourist destinations in Spain, and also in the fact that Switzerland, depicted in various popular yesteryear Indian films (movies), remains a popular tourist destination for Indians to date [7].

The Hollywood co-actor network is a social network that has invited a fair amount of interest in the past [8], studies being conducted using relational dependency network analysis, Layered Label Propagation algorithm and PageRank algorithm [9, 10]. In comparison, its much larger counterpart in India has been largely ignored. Flourishing with a $9\%$ growth from 2009 to 2010 [7] and a further $11.5\%$ growth from 2010 to 2011 [11], it is an industry that sees blazingly fast growth, leading us to expect drastic changes in small time frames. We study the Bollywood industry because it provides a fair ground to capture the temporal changes in a network owing to its rapidly changing character. Using data from the past 100

years, we construct a network for every five year period. The nodes can be classified into the three distinct categories : 1) lead male actors, 2) lead female actors and 3) supporting actors. We analyze the structural properties of this network and further study its spectral properties using the random matrix theory (RMT).

Though originally rooted in nuclear physics [14], RMT has found widespread applications in different real systems such as the stock-market indices, atmosphere, human EEG, large relay networks, biological networks and various other model networks. Under the framework of RMT, such systems and networks follow the universal Gaussian orthogonal ensemble (GOE) statistics. Though there exist other universality classes such as Gaussian unitary ensemble and Gaussian symplectic ensemble [13], which have also been extensively investigated in RMT literature, we focus only on GOE statistics as spectra of various networks have been shown to rest with this universality class [14–16]. The universality means that universal spectral behaviors, such as statistics of nearest neighbor spacing distribution (NNSD) are not only confined to random matrices but get extended to other systems. A wide variety of complex systems fall under this class, i.e. their spectra follow GOE statistics ( [17] and references therein).

# 2 Materials and Methods

## 2.1 Construction of Bollywood networks

We collect all Bollywood data primarily from the movie repository website *www.bollywoodhungama.com* and henceforth from *www.imdb.com* and *www.fridayrelease.com* (now renamed as *www.bollywoodmdb.com*) and we generate no additional data. The website *www.bollywoodhungama.com* previously known as *www.Indiafm.com*, is a reputed Bollywood entertainment website, owned by Hungama Digital Media Entertainment, which acquired Bollywood portal in 2000. We use Python code to extract names of all the movies and their corresponding information for a massive period of hundred years spanning from 1913 to 2012. Initially we document the names of all films as per their chronological sequence (latest to oldest) from the websites by incorporating the desired URL [18] in the code along with a built-in string function which takes the page numbers (932 pages in "Released before 2012" category and 24 pages in "Released in 2012" category) as input. Each film of every page bears a unique cast ID in the website, navigating to which via "Movie Info" provides us complete information about the film. In the Python code, we store the unique cast IDs of films in a temporary variable and retrieve relevant information using appropriate keywords from the respective html page. We also manually browse through other aforementioned websites in order to collect any year-wise missing data, if any. Thus we get the data in terms of names of the movies and names of the actors for 100 years. We then merge the data from all the websites and omit repetitions. A total of 8931 movies have been documented so far in Bollywood from 1913 till 2012. Harvesting the complete data took approximately 2000 hours of work over a 4-month period, which includes manual verification, formatting, removal of typos and compilation of the data. Considering the rapidly changing nature of the Bollywood network, we assort the curated massive Bollywood data in to 20 datasets each containing movie data for five-year window periods, as this is an apt time frame within which the network constructed is large enough to study the important network properties, and is not too large to miss any crucial evolutionary information. Since the number of movies and their actors in the time span 1913-1932 were scanty and could not have yielded any significant statistics, we merge the 1913-1932 datasets and present as a single dataset 1928-1932.

We create database of all actors who had appeared in the Bollywood film industry ever since its inception in five-year window periods, as mentioned in the previous version of the manuscript, by extracting them from the movie information using Python algorithm and we assign a unique ID number to each actor in every span which we preserve throughout our analysis. We take care of ambiguities in spellings of names of actors presented in different websites by extensive thorough manual search and cross-checking to avoid overlapping of information and duplication of node identities while constructing networks. Tracking by

their unique ID numbers assigned by us, we create a co-actor database for each span where every pair of actors who had co-acted in a movie within those five years are documented. We then construct an adjacency list of all available combinations of co-actors. Treating every actor as a node and every co-actor association as a connection, we create a co-actor network of the largest connected component for every span.

We pick the actors appearing as the protagonist (occupant of the first position) in the movie star cast list from the movie star cast database created by us and observe that they incidentally are male actors in almost all movies with some rare exceptions. On extensive manual search based on popularity, award nominations we find that those male actors appear as a lead in the respective movies which made our attempt to extract lead male actors even easier. We could very well define the lead male actor as the protagonist in the star cast of at least five films in consecutive five-year spans and extract them from the movie star cast list using Python code while we were unable to find any proper definition for lead female actors as the second position of the movie star cast list is alternately occupied by either female actors or supporting actors, making it difficult to extract them only based on the network data as described. Hence we handpick the lead female actors from the movie star cast database for all the spans based on their popularity, award nominations and create their database.

## 2.2 Assimilation of Filmfare awards data

We consider Filmfare award nominations as the best means to assess the success rates of all lead actors of Bollywood and distinguish the lead female actors from the rest. Filmfare awards were first introduced by the The Times Group [19] after the Central Board of Film Certification (CBFC) was founded by Indian central government in 1952 to secure the identity of Indian culture. The reason behind choosing Filmfare Awards amongst all other awards in our analysis is that it is voted both by the public and a committee of experts, thus gaining more acceptance over the years. Instead of the awards bagged we rather take into account the award nominations in order to avoid the interplay of some kind of bias affecting the decision of the CBFC committee in selecting the winner. By manual navigation through every year of Filmfare awards available on the web, we create a database of all categories of Filmfare awards and extract their respective nominees chronologically from the html pages using Python codes. Henceforth we use C++ codes to count the number of times every actor is nominated in each five-year span. Thus we obtain a complete list of all actors in each span along with their number of Filmfare nominations.

## 2.3 Structural attributes of Bollywood networks

Considering $p_k$ to be the fraction of vertices with the degree k, the degree distribution of the constructed networks is plotted with $p_k$. It has been sufficiently proven that the degree distribution of real world networks are not random, most of them having a long right tail corresponding to values that are far above the mean [8].

We define the betweenness centrality of a node i, as the fraction of shortest paths between node pairs that pass through the said node of interest [9].

$$x_i = \sum_{st} \frac{n_{st}^i}{g_{st}} \tag{1}$$

where $n_{st}^i$ is the number of geodesic paths from $s$ to $t$ that passes through $i$ and $g_{st}$ is the total number of geodesic paths from $s$ to $t$.

## 2.4 Measures used for success appraisal

In the current work, the concept of a payoff has been borrowed from the field of management [21], and adapted to suit the Bollywood network analysis. Payoff has elucidated the success of the center and non-

center agents in a unique efficient star network [22]. We use an improvised version of payoff as a means to assess success rates of the nodes in Bollywood. For the purpose of devising net payoff ($P_i$), we study the datasets two at a time (accounting for ten years) and use the following definition:

$$P_i = \frac{1}{\Delta d_i} + \langle \sin(\pi d_n) \rangle + \langle \sum_j w_j \left( \frac{1}{n_i} + \frac{1}{n_j} + \frac{1}{n_i n_j} \right) \rangle \tag{2}$$

where, $\Delta d_i$ is the change in degree of a particular node $i$ in two consecutive spans. $d_n$ is its normalized degree in a particular span given as $d_n = (\frac{d_i - d_{min}}{d_{max} - d_{min}})$ with $d_i$ being the degree of the node i and $d_{max}$ and $d_{min}$ being the maximum and minimum degree in that particular span, respectively. The third term sums over all nodes j that node i has worked with where $n_i$ and $n_j$ are the number of movies that the node i and j has worked in respectively and $w_j$ the number of times the node $j$ has worked with the node $i$ in the considered time window. The averages denoted in the net payoff (Eq. 2) refer to the values averaged over the two consecutive datasets. Based on the values of $P_i$, the actors of every set studied were ranked and lists made.

Due to the absence of a unifying framework that can be used to evaluate the success of films and their actors in the years before the inception of Filmfare Awards in 1954, we restrict our analysis on assessment of success to the time periods spanning from 1954 and onwards. In order to adumbrate the success of actors in the industry, we define overlap as the intersection of sets of co-actors that an actor has worked with, in two consecutive time frames.

## 2.5 Spectral analyses

The random matrix studies of eigenvalue spectra consider two properties: (1) global properties such as spectral distribution of eigenvalues $\rho(\lambda)$, and (2) local properties such as eigenvalue fluctuations around $\rho(\lambda)$. Eigenvalue fluctuations is the most popular one in RMT and is generally obtained from the NNSD of eigenvalues. We denote the eigenvalues of a network by $\lambda_i = 1, \ldots, N$ and $\lambda_1 > \lambda_2 > \lambda_3 > \ldots > \lambda_N$. In order to get universal properties of the fluctuations of eigenvalues, it is customary in RMT to unfold the eigenvalues by a transformation $\bar{\lambda}_i = \bar{N}(\lambda_i)$, where $\bar{N}$ is average integrated eigenvalue density. Since we do not have any analytical form for $N$, we numerically unfold the spectrum by polynomial curve fitting [14]. After unfolding, average spacings are unity, independent of the system. Using the unfolded spectra, spacings are calculated as $s^{(i)} = \bar{\lambda}_{i+1} - \bar{\lambda}_i$.

The NNSD is given by

$$P(s) = \frac{\pi}{2} s \exp\left( -\frac{\pi s^2}{4} \right). \tag{3}$$

For intermediate cases, the spacing distribution is described by Brody distribution as

$$P_\beta(s) = A s^\beta \exp\left( -\alpha s^{\beta+1} \right) \tag{4}$$

where $A$ and $\alpha$ are determined by the parameter $\beta$ as follows:

$$A = (1+\beta)\alpha, \;\; \alpha = \left[ \Gamma\left( \frac{\beta+2}{\beta+1} \right) \right]^{\beta+1}$$

This is a semi-empirical formula characterized by parameter $\beta$. As $\beta$ goes from 0 to 1, the Brody distribution smoothly changes from Poisson to GOE. Fitting spacing distributions of different networks with the Brody distribution $P_\beta(s)$ gives an estimation of $\beta$, and consequently identifies whether the spacing distribution of a given network is Poisson, GOE, or the intermediate of the two [15].

The NNSD accounts for the short range correlations in the eigenvalues. We probe for the long range correlations in eigenvalues using $\Delta_3(L)$ statistics which measures the least-square deviation of the spectral staircase function representing average integrated eigenvalue density $N(\bar{\lambda})$ from the best fitted straight line for a finite interval of length $L$ of the spectrum and is given by

$$\Delta_3(L;x) = \frac{1}{L} min_{a,b} \int_x^{x+L} [N(\overline{\lambda}) - a\overline{\lambda} - b]^2 d\overline{\lambda} \tag{5}$$

where $a$ and $b$ are regression coefficients obtained after least square fit. Average over several choices of x gives the spectral rigidity, the $\Delta_3(L)$. In case of GOE statistics, the $\Delta_3(L)$ depends logarithmically on L, i.e.

$$\Delta_3(L) \backsim \frac{1}{\pi^2} \ln L \tag{6}$$

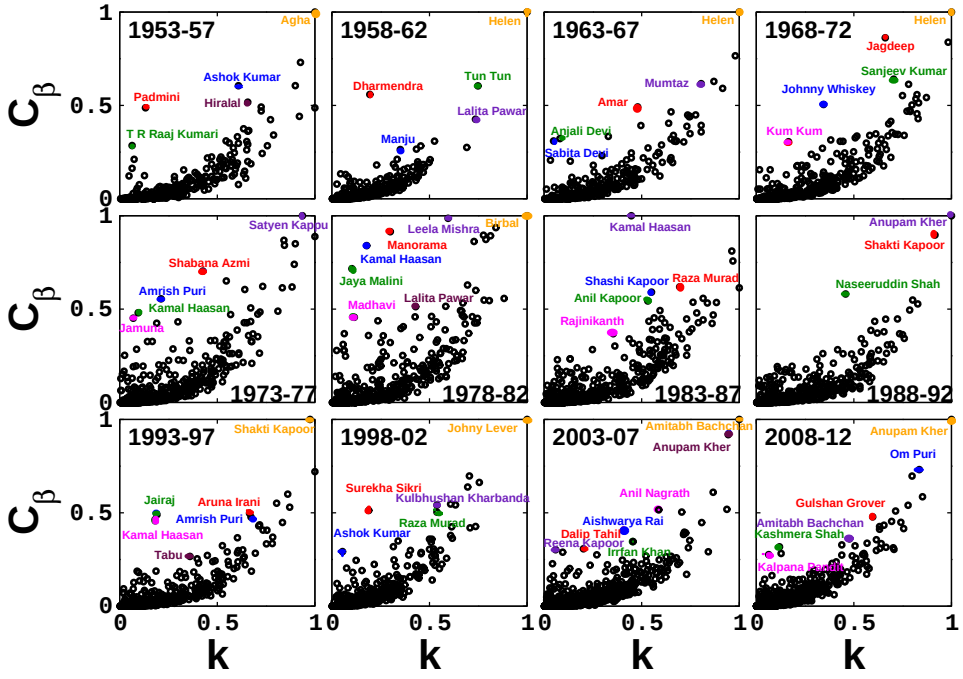



Figure 1: (Color online) Plots of normalized betweenness centrality ($C_\beta$) against normalized degrees ($k$) of Bollywood actors over 1953-2012. Actors and their corresponding betweenness centrality are represented in same color.

# 3 Results and Discussion

## 3.1 Structural properties of Bollywood networks

The degree distribution of the Bollywood networks follow power law, as expected based on the studies of other real world networks [8]. But an observation that defies intuition is that the most important nodes of the industry, acknowledged as the lead male actors, do not form the hubs of the constructed network, but instead have a moderate degree and also maintain it along sets of data that were studied (SI Tables 1-6). Considering the network on an evolutionary scale, this is a property that gains more prominence during the later sets of the data, while the network maintains power law throughout the entire timespan (SI Fig. 1). The prominent supporting actors of the era form the hubs of the industry in respective time frames. This counterintuitive

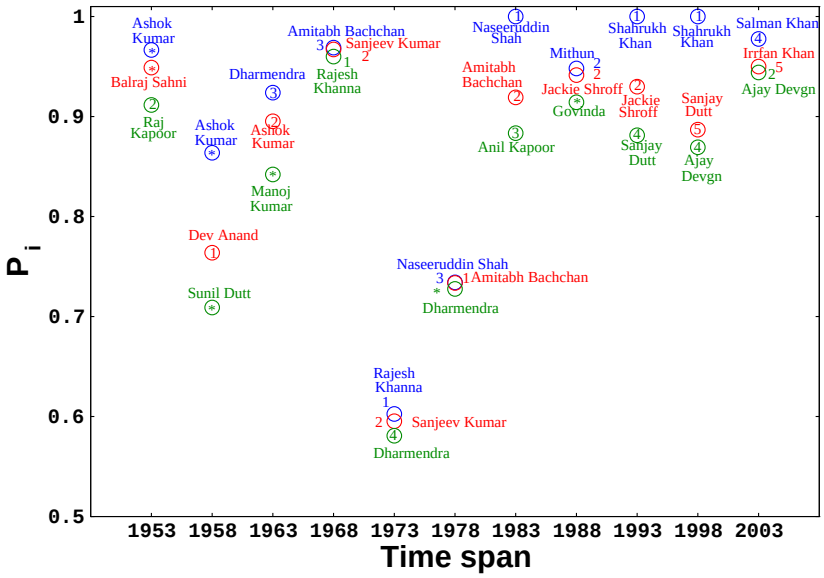



Figure 2: (Color online) Net payoff ($P_i$) of top three lead male actors. in each time span plotted against the respective time frames. They are ranked (as 1, 2 and so on) based on their number of Filmfare award nominations. '*' denotes no Filmfare award nominations. Actors and their corresponding rankings are represented in same color.

nature of the above observation can be explained by the fact that these actors collaborate with more nodes and take on more projects in a given time period. Hence they can be said to be instrumental in establishing connections in the network. The scale-free behavior of the Bollywood industry can be elucidated by the fact that newcomers in the industry in general aspire to act with the lead actors of the era, who intuitively form associations with high degree nodes, thus illustrating the preferential attachment property prevalent in Bollywood networks [8].

## 3.2 Success appraisal of Bollywood actors

By virtue of the sinusoidal function used in (Eq. 2), the nodes with a moderate degree lead the net payoff list with both low degree and high degree nodes trailing behind. The inverse of the change in degree favors nodes that preserve their degree over the years hence giving a higher net-payoff to actors who preserve their degrees over the various datasets.

Successful supporting actors, although bear a high degree, appear quite high in the scale of $P_i$ because they have relatively higher values of $\langle p_i \rangle$. Though interplay of various contrasting factors influence the appearance of lead male actors in $P_i$ list, they appear high in absolute scale of $P_i$ in all the sets under consideration except the ones corresponding to 1973-77 and 1978-82. Three of the top five Filmfare award nominees in lead male actor category appear as top three lead male actors in $P_i$ list in respective time frames (Fig. 2 and SI Tables 1-6). This observation is more pronounced in case of the lead female actors. As observed in Fig. 3 and SI Tables 7-12, the three lead female actors having secured the maximum number of Filmfare award nominations in a particular span of time, appear as the leading nodes in their respective $P_i$ list, a trait that is more consistent in the more recent datasets. From the above analysis based on payoff it is supposed that possessing moderate degree and maintaining it are properties followed by the nodes that stand successful in Bollywood industry and can be contemplated as keys to success. Succeeding the economic liberalization in 1991, the inclusion of diverse socio-political-economic issues in mainstream Bollywood movies found favor with the audience [24]. At around this period, Hollywood started gaining popularity among the Indian population owing to the advent of private movie channels and the internet.

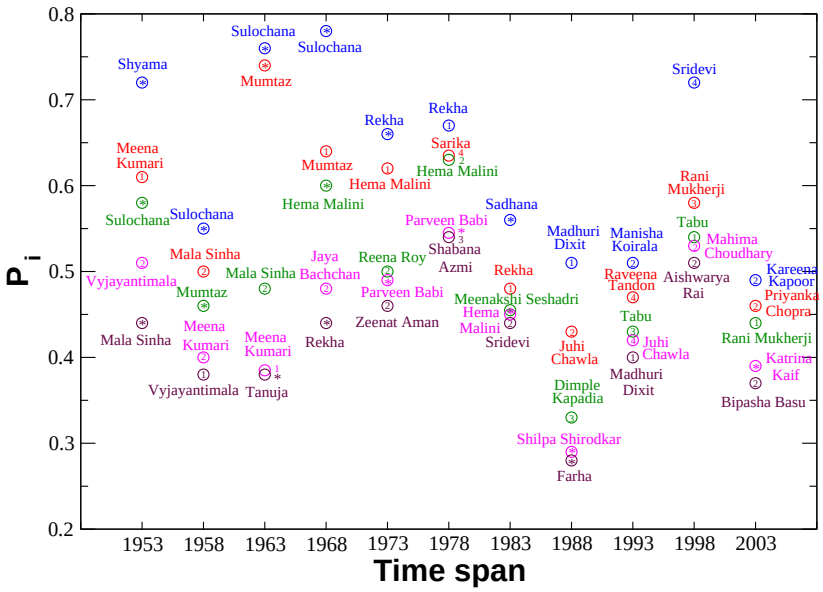



Figure 3: (Color online) Net payoff ($P_i$) of top five lead female actors in each time span plotted against the respective time frames. They are ranked (as 1, 2 and so on) based on their number of Filmfare award nominations. '*' denotes no Filmfare award nominations. Actors and their corresponding rankings are represented in same color.

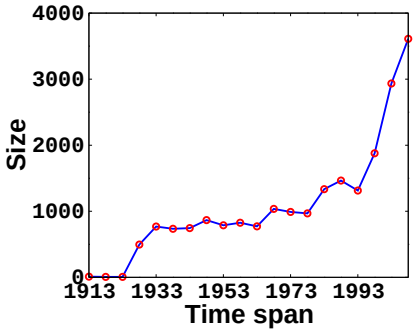



Figure 4: (Color online) Evolution of Bollywood network size over 1913-2012.

These factors coupled together affected the structure of the network, which might be the underlying reason behind the observed variations in the network properties, pre, post and during liberalization. A steep rise in the Bollywood network size 1993 onwards (Fig. 4) might be one of the manifestations of this shift in economic policies. The status of an 'industry' being conferred upon Bollywood in 1998 might be a result of this increased size of the network [25]. The comparatively larger shift of the network properties with the advent of liberalization as opposed to that caused by the introduction of the Filmfare awards in 1954, can lead us to conclude that mainstream Bollywood is largely driven by economic concerns rather than artistic ones. The number of times an actor is nominated for the Filmfare awards while they remain a lead actor, when plotted with their overlap (as defined before), shows that 22 among the 25 actors exhibit an approximate direct proportionality (Fig. 5) emphasizing on the importance of winning combinations. Overlap being one of the probable factors deciding the success of a node might explain the reason for the formation of social groups, and co-operation among them in the society [26].

High degree nodes indubitably have high betweenness centrality. Actors with high betweenness centrality seem to have a relatively larger span in the industry even if their popularity levels, measured as the number of Filmfare award nominations, is not markedly high. Nodes with the highest betweenness centrality of all datasets are found to be male actors (except Helen), whether lead or supporting, adumbrating the gender disparity in Bollywood. Incidentally, few of the nodes bearing moderate and low degree also

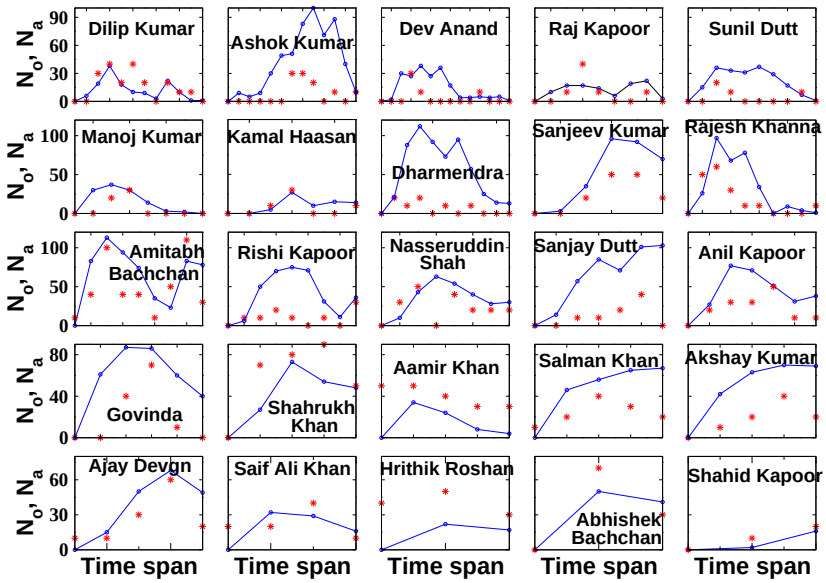



Figure 5: (Color online) Plots of individual overlaps $N_o$ (represented by $\bullet$) of lead male actors and their Filmfare award nominations $N_a$ (represented by $*$) against their respective time spans. Time span here represents respective individual spans of lead male actors in Bollywood industry, for example Dilip Kumar had a long span stretching between 1943 and 1998 whereas Hrithik Roshan has a short spell 1998 onwards.

exhibit high betweenness centrality and also have a long span in the Bollywood industry (Fig. 1; SI Fig. 2 and Table 2). This indicates that actors exhibiting mobility between diverse Bollywood circles seem to have an advantage of a long span, though we are far from concluding that this is the only factor affecting the life span of a node. There exist examples from social and biological systems which also support the importance of cooperation and mobility [27].

## 3.3 Spectral analyses of Bollywood networks

The spectral density, $\rho(\lambda)$ of the connectivity matrix of Bollywood networks exhibit a triangular distribution (SI Fig. 3 and discussion in [28]), hence providing evidence supporting its scale-free nature [29]. The eigenvalue distribution of the Bollywood networks show a high degeneracy at $-1$, deviating from the commonly observed degeneracy at $0$ in most of the real world networks studied (for example, biological networks [14]) . This degeneracy at $-1$ can be attributed to the presence of clique structures in the network [30]. Presence of dead-end vertices in spectrum and motif joining or duplication have been used as plausible explanations to widespread degeneracy at $0$ observed in biological networks [31]. Factors affecting a social network are vastly different from those affecting a biological network, hence making the nature of their spectra varied. Owing to a relatively smaller number of nodes in the networks constructed for the periods 1913-17, 1918-22 and 1923-27, a bulk does not appear in their eigenvalue distributions. The distributions corresponding to the datasets of 1928-57, 1983-87 and 2003-12 very clearly show the presence of a few eigenvalues outside the bulk (SI Fig. 4 and Fig. 6), which is formed by the rest of the eigenvalues. While the largest eigenvalue is distinctly separated from the bulk, which is a well-known spectral feature of an undirected network [9], existence of other eigenvalues outside the bulk probably indicate the existence of distinct Bollywood guilds [32] further portending an evolving network structure.

The spectral data as well as the data regarding the betweenness centrality of the networks, corresponding to the time periods after 1998-02, suggest that there has been a drastic change in the underlying network structure since then. This marked change in the more recent datasets in comparison to the older ones, is clearly illustrated by the presence of several eigenvalues outside the bulk (Fig. 6), and the presence

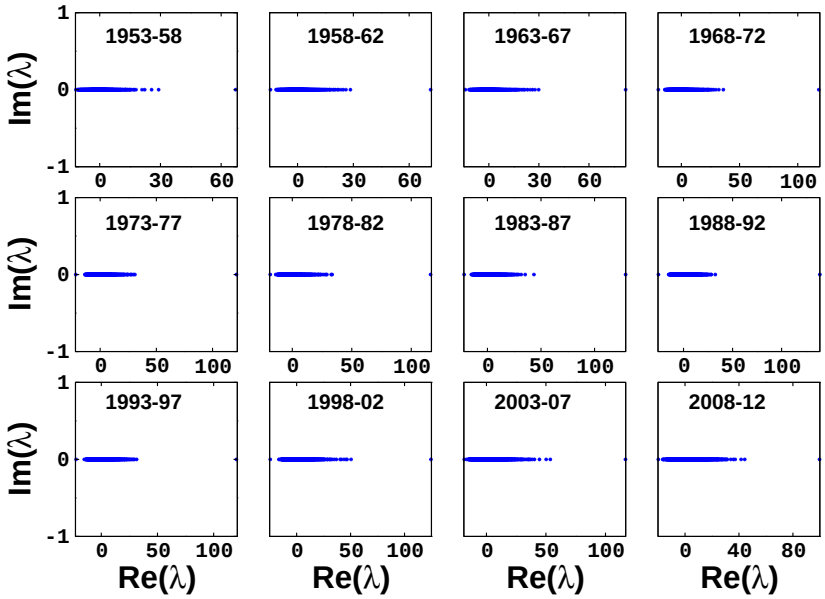



Figure 6: (Color online) Separation of lone eigenvalues from bulk of eigenvalues in Bollywood datasets spanning between 1953-2012.

of a lesser number of low degree nodes with a high betweenness centrality(Fig. 1). This indicates that the community structures in the Bollywood network have gotten more inter-interconnected post 1998-02, leading the authors of this paper to conclude that Bollywood is becoming increasingly systematic with time.

We fit the NNSD of Bollywood networks by the Brody distribution (Eq. 4) and find that the value of $\beta$ comes out to be close to $1$ for all the datasets. This implies that the NNSD of Bollywood datasets follow GOE statistics of RMT (Eq. 3 and SI Fig. 5) bringing Bollywood networks under the universality class of RMT [15, 17]. To examine the long range correlations, we calculate spectral rigidity via the $\Delta_3(L)$ statistics of RMT using Eq. 5 by taking same unfolded eigenvalues of different datasets as used for the NNSD calculations. The value of $L$ for which the $\Delta_3(L)$ statistics follows RMT prediction (Eq. 6) is given in the Table 1 and the detailed plots are deferred to [28] as SI Fig. 6. The $\Delta_3(L)$ statistics which provides a measure of randomness in networks [16] clearly indicate that the dataset corresponding to the 1963-67 timespan has the most random underlying network structure when compared with the other datasets. This notable feature of this timespan can probably be attributed to the consecutive wars that India was a part of in the years 1962 and 1965, which in turn lead to an extreme economic crisis in the country. As shown by the decreasing value of L since 1933, the networks have a trend of diminishing randomness.The dataset corresponding to 1948-52 witnessed a breach from this trend, probably due to the drastic political and financial changes post Indian Independence in 1947. One of the most crucial points exhibited in the analysis based on eigenvalue distribution and betweenness centrality is that, before the year 1998 the structure of the networks had either well segregated clusters or extreme random interactions, while post 1998 the structures seem to maintain a fairly consistent randomness (randomness measured by the value of L).

# 4 Conclusions

Although Bollywood networks for different spans demonstrate varying amounts of randomness as suggested by the changing values of L in the $\Delta_3(L)$ statistics, observation of universal GOE statistics of the NNSD puts forward the evidence to show that a sufficient amount of randomness is possessed by all the sets. The efficiency of many real world systems such as the financial markets, the climatic system, neuronal systems etc, has been aided by their stochastic nature which leads to randomness [33]. Bollywood network also

Table 1: **Properties of Bollywood network of each 5 years block datasets.**

| Time span | $N$ | $\langle k \rangle$ | $N_eff$ | $L$ | % $\Delta_3(L)$ |
|---|---|---|---|---|---|
| 1928-32 | 496 | 9.46 | 162 | 8 | 4.93 |
| 1933-37 | 769 | 10.7 | 246 | 6 | 2.43 |
| 1938-42 | 735 | 13.3 | 248 | 5 | 2.02 |
| 1943-47 | 745 | 12.6 | 276 | 5 | 1.81 |
| 1948-52 | 866 | 17.5 | 291 | 8 | 2.75 |
| 1953-57 | 788 | 25.9 | 272 | - | - |
| 1958-62 | 827 | 29.9 | 313 | - | - |
| 1963-67 | 772 | 35.2 | 308 | 19 | 6.16 |
| 1968-72 | 1036 | 47.0 | 416 | - | - |
| 1973-77 | 990 | 47.5 | 383 | 14 | 3.65 |
| 1978-82 | 968 | 45.1 | 370 | 16 | 4.32 |
| 1983-87 | 1335 | 44.6 | 480 | 19 | 3.95 |
| 1988-92 | 1465 | 44.9 | 546 | 24 | 4.39 |
| 1993-97 | 1314 | 42.2 | 504 | 12 | 2.38 |
| 1998-02 | 1878 | 46.3 | 686 | 14 | 2.04 |
| 2003-07 | 2935 | 37.0 | 973 | 17 | 1.74 |
| 2008-12 | 3611 | 30.3 | 1164 | 17 | 1.46 |

$N$ and $\langle k \rangle$ respectively denote size and average degree of network. $N_{eff}$ and $L$ are the effective dimension of non-degenerate eigenvalues less than $-1$ and the length of the spectrum up to which spectra follow RMT. % The $\Delta_3(L)$ represents the extent of $L$ 2 which spectra follow GOE statistics, expressed in percentage terms. '-' denotes the spectra which do not follow RMT.

provides an example to aid this relationship, as the industry has survived various valleys and crests since its inception, including in times of dire socio-economic crisis [34]. The extensive analyses of Bollywood data on the one hand reveals its influence on the decisions and preferences of the mass, while on the other it unravels the prevailing gender disparity [35, 36] thus acting as a reflection of the society. Furthermore, it helps us deduce that cooperation among the nodes leads to combinations that become formulaic for successful ventures. It also seems to further propagate the idea suggesting that a combination of organization and randomness in the network structure supports the sustenance of the represented network. We believe that the analysis of the Bollywood network as carried out in this work can be extrapolated to study the predictability of success and the ingredients that are necessary for the robustness of other social collaboration networks [37] and organizations [38].

# 5 Acknowledgments

AM acknowledges IIT Indore for providing a conducive environment for carrying out her internship. We are grateful to Arul Lakshminarayan (IITM) for time to time fruitful discussions on random matrix aspects and Dima Shepelyansky (Université Paul Sabatier) for useful suggestions. We are thankful to the Complex Systems Lab members, Ankit Agrawal and Aradhana Singh for helping with data download and discussions.

# References


[1] Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Review of Modern Physics 74: 47-97.

[2] Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. Science 323: 892-895.

[3] Carrington PJ, Scott J, Wasserman S (2005) Models and methods in social network analysis. Cambridge University Press, New York, 1st edition, 1-44 pp.

[4] Song C, Qu Z, Blumm N, Barabási AL (2010) Limits of predictability in human mobility. Science 327: 1018-1021.

[5] KPMG India, Confederation of Indian Industry (2005) A CII-KPMG Report: Indian Entertainment Industry Focus 2010-Dreams to Reality. KPMG India and Confederation of Indian Industry.

[6] Bose M (2007) Bollywood: A History, 1st edition. New Delhi: Rakmo Press. pp. 37362.

[7] KPMG India, Federation of Indian Chambers of Commerce and Industry(2011) FICCI-KPMG Indian Media and Entertainment Industry Report: Hitting the High Notes. KPMG India and Federation of Indian Chambers of Commerce and Industry.

[8] Martino F, Spoto A (2006) Social network analysis: A brief theoretical review and further perspectives in the study of information technology. PsychNology Journal 4: 53-86.

[9] Cattani G, Ferriani S (2006) A core/periphery perspective on individual creative performance: Social networks and cinematic achievements in the hollywood film industry. Organization Science 4: 53-86.

[10] Boldi P, Rosa M, Vigna S (2011) Robustness of social networks: Comparative results based on distance distributions. Social Informatics 6984: 8-21.

[11] KPMG India, Federation of Indian Chambers of Commerce and Industry(2012) FICCI-KPMG Indian Media and Entertainment Industry Report: Digital Dawn The metamorphosis begins. KPMG India and Federation of Indian Chambers of Commerce and Industry.

[12] Mehta ML (1991) Random Matrices, 2nd edition. New York: Academic Press.

[13] Akemann G, Baik J, Francesco PD (2011) The Oxford Handbook of Random Matrix Theory, 1st edition. Oxford: Oxford University Press.

[14] Jalan S, Ung CY, Bhojwani J, Li B, Zhang L, et al. (2012) Spectral analysis of gene co-expression network of zebrafish. Europhysics Letters 99: e48004(1-6).

[15] Jalan S, Bandyopadhyay JN (2007) Random matrix analysis of complex networks. Physical Review E 76: e046107(1-7).

[16] Jalan S, Bandyopadhyay JN (2009) Randomness of random networks: A random matrix analysis. Europhysics Letters 87: e48010(1-5).

[17] Guhr T, M-Groeling A, Weidenm$\ddot{u}$ller HA (1998) Random-matrix theories in quantum physics: common concepts. Physics Reports 299: 189-425.

[18] Bollywoodhungama website. Available: http://akm-www.bollywoodhungama.com/movies/list/sort/Released be listing/page/ and http://akm-www.bollywoodhungama.com/movies/list/sort/Released in 2012/ char/ALL/type/listing/page/. Accessed 2013 August 10.

[19] Filmfare website. Available: http://www.filmfare.com. Accessed 2013 August 10.

[20] Newman MEJ (2003) The structure and function of complex networks. SIAM Review 45: 167-256.

[21] Jackson MO, Wolinsky A (1996) A strategic model of social and economic networks. Journal of Economic Theory 71: 44-74.

[22] Watts A (2001) A dynamic model of network formation. Games and Economic Behavior 34: 331-341.

[23] Brody TA (1973) Statistical measure for repulsion of energy-levels. Lett Nuovo Cimento 7: 482- 484.

[24] University of Chicago (2006) Task Force Report: Economic reforms in India. Chicago, IL: University of Chicago.

[25] Ray R (2012) Wither slumdog millionaire: Indias liberalization and development themes in bollywood films. 17th International Business Research Conference, Toronto, Canada.

[26] Pacheco JM, Santos FC, Chalub FACC (2006) Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. PLoS Computational Biology 2: 1634-1638.

[27] Súarez YR, Júnior MP, Catella AC (2004) Factors regulating diversity and abundance of fish communities in pantanal lagoons, brazil. Fisheries Management and Ecology 11: 4550.

[28] Supporting Information.

[29] de Aguiar MAM, Bar-Yam Y (2005) Spectral analysis and the dynamic response of complex networks. Physical Review E 71: e016106(15).

[30] Mieghem PV (2011) Graph Spectra for Complex Networks, 1st edition. New York: Cambridge University Press. pp. 11345.

[31] Dorogovtsev SN, Goltsev AV, Mendes JFF, Samukhin AN (2003) Spectra of complex networks. Physical Review E 68: e046109(110).

[32] Chauhan S, Girvan M, Ott E (2009) Spectral properties of networks with community structure. Physical Review E 80: e056114(110).

[33] Gammaitoni L, Hanggi P, Jung P, Marchesoni F (1998) Stochastic resonance. Review of Modern Physics 70: 223287.

[34] Research Unit (LARRDIS), Rajya Sabha Secretariat, New Delhi (2009) Global economic crisis and its impact on India. New Delhi, India.

[35] Das D, Pathak M (2012) Gender equality: A core concept of socio-economic development in india. Asian Journal of Social Sciences and Humanities 1: 257 264.

[36] Kristof ND, WuDunn S (2009) Half the Sky: Turning Oppression into Opportunity for Women Worldwide, 1st edition. New York: Vintage Publishing. pp. 1294.

[37] Guimera R, Uzzi B, Spiro J, Amaral LAN (2005) Team assembly mechanisms determine collaboration network structure and team performance. Science 308: 697702.

[38] Tichy NM, Tushman ML, Fombrun C (1979) Social network analysis for organizations. Academy of Management Review 4: 507519.

# Supporting Information

## Uncovering randomness and success in society

**Sarika Jalan, Camellia Sarkar, Anagha Madhusudanan, Sanjiv Kumar Dwivedi**

## 1. Methods

Study of society and its movement has traditionally involved obtaining data from representative populations through field studies and extrapolating the obtained results through approximations [1]. These methods of data collection provide, in the first place incomplete data and secondly, data that is prone to errors that would drastically skew the results obtained by the physicists' method of studying them. Movie actors networks analyses became a lucrative means for assessing society as the data obtained is to a satisfiable extent accurate and free from approximations and bias.

Although individual endowments (income) should rationally be the apt discriminating factor for distinguishing lead actors from the supporting ones, it is quite cumbersome to retrieve relevant data due to lack of reliable sources meant for the same. The variable nature of the data adds to its impediment. We define lead male actors based on the number of times they top the starcast list in consecutive spans while defining lead female actors still remains an agony even after a century of cinematic heritage (discussed in sufficient detail in the main article). Although movies like Fashion, Page 3, Chandni Baar, Kahaani, Heroine portrays the never ending struggle of women in society, the basis of their struggles have undoubtedly changed over the years. While Mother India (1957) depicts the struggle for existence, a struggle to combat poverty, Fashion (2008) depicts a struggle for fame, a struggle for passion, a struggle for touching dreams, but not a struggle for existence. This reflects a gradual change in the outlook of the society towards women.

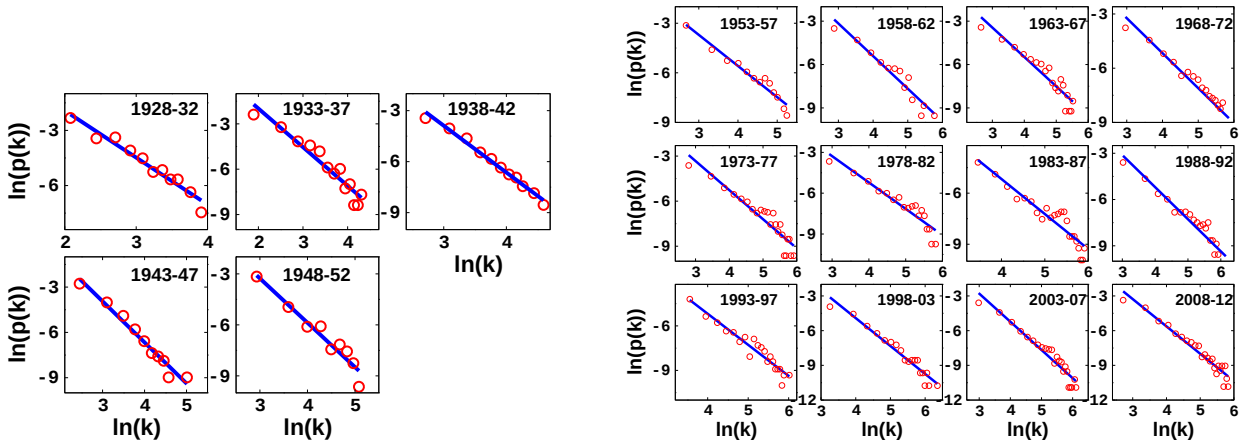
In order to assess success of all actors in Bollywood industry, the Filmfare Awards were introduced for rewarding both artistic and technical excellence of professionals in the Hindi language film industry of India. The National Film Awards were also introduced in 1954 but gained less popularity as compared to Filmfare as they are decided by a panel appointed by Indian Government and do not authentically reflect the choice of the global audience. The Filmfare Awards, in contrast, are voted for by both the public and a committee of experts thus gaining more acceptance over the years.

### 1.1 A brief review of Hollywood networks

The collaboration graph of film actors were shown to be small-world networks [2] and their properties were studied using random graph theory [3]. Relational dependency network analysis has been performed on Hollywood datasets obtained from IMDB which identify and exploit cyclic relational dependencies to achieve significant performance gains [4]. Hollywood datasets were deployed for implementation of the Layered Label Propagation algorithm, meant to reorder very large graphs [5] and the PageRank algorithm to uncover the relative importance of a node in a graph [6]. Professional links between movie actors was used as a means to fit the predictions of a continuum theory to probe for the existence of two regimes, the scale-free and the exponential regime [7].

## 1.2 Structural Analyses

### 1.2.1 Degree Distribution

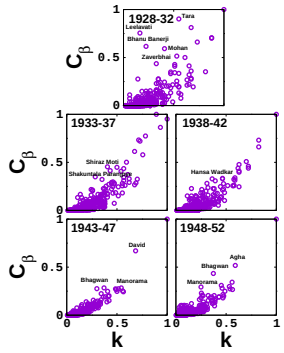



SI Figure 1: Degree distribution of the Bollywood networks over 1913-2012. Due to scarcity of actors in 1913-1927, all nodes appearing in 1913-27 have been merged and included in 1928-32.

Degree of a node can be defined as the number of nodes that are linked to the said node. Degree distribution is the plot of the degree versus the number of nodes with the particular degree. SI Fig. 1 plots degree distribution of Bollywood networks.

### 1.2.2 Betweenness Centrality

The supporting actors have been observed to have high betweenness centrality. Nodes having higher degree would naturally be coming into shortest path between pair of nodes, and hence would have high betweenness centrality. Fig. 4 of main article and SI Fig.2 has highest $C_\beta$ corresponding to node possessing highest degree. The fact that larger degree in any of the sets in 1928-2012 are possessed by supporting actors, and it is somewhat established that supporting actors have longer life span than lead male actor and lead female actors, makes the positive correlation between degree and life span quite obvious. But some of the low




SI Figure 2: Plots of normalized betweenness centrality ($C_\beta$) against normalized degrees ($k$) of Bollywood actors over 1913-1952.

degree nodes are also seen to have high betweenness centrality. Either they are supporting actors which again comply with the earlier argument for their larger life span, or if they are lead male actors then also they show accredited life span. For example, in 1958-62 dataset, Dharmendra having low degree distinctly

appears in the high betweenness centrality region and has a remarkably long span (1953-2012) in the industry. Few other prominent actors who have been seen to follow this trend are Kamal Haasan (1958-2012), Nasseruddin Shah (1973-2012), Rajinikanth (1973-2012), Anil Kapoor (1978-2012). These examples are taken for those who are clearly depicting high betweenness centrality than rest of the nodes around them. Various female actors having low degree also fall in high betweenness centrality region and have long span. Padmini (1948-77) and Rajinikanth (1973-2012) are Tamil actors who have been observed in high betweenness centrality region bridging the gap between communities of Bollywood and Kollywood (Table 2).

Table 2: List of prominent actors who appear high in betweenness centrality zone

| Names of actors | Span | Recognition |
|---|---|---|
| Agha | 1937-1989 | Known for comic roles, won Filmfare Best Supporting Actor Award (1960) |
| Ashok Kumar | 1936-1993 | An iconic figure in Indian cinema popularly known as "Dadamoni" who is also a painter, homeopath, astrologer, boxer, chess player, singer ; confered with honors like Dadasaheb Phalke award (1988) and Padma Bhushan (1998), Filmfare Lifetime Achievement Award (1995), Sangeet Natak Akademi Award (1959), National Film Awards for Best Actor (1969), Filmfare awards (1962, 1966, 1969) |
| Padmini | 1948-1994 | An elegant **Tamil** dancer who was also featured in several Hindi films; won Filmfare Award for Best Supporting Actress (1966) |
| Hiralal | 1928-1995 | A prominent supporting actor having a long span in industry |
| T R Rajakumari | 1936-1955 | Originally a Tamil film actress, Carnatic singer and dancer also acted in many Bollywood films |
| Helen | 1951-2012 | An Indian film actress and one of the most popular dancers of all times; has bagged Padma Shri (2009), Filmfare Best Supporting Actress Award (1979), Filmfare Lifetime Achievement Award (1998) |
| Tun Tun | 1946-1990 | A highly rated playback singer who later became a permanent comic relief in numerous Bollywood films. |
| Dharmendra | 1960-2012 | Often referred to as the "He-Man", he has won Padma Bhushan (2012), Filmfare Lifetime Achievement award (1997), Filmfare Best Actor awards (1967, 1972, 1974, 1975), the Living Legend award (FICCI) and many more |
| Lalita Pawar | 1928-1997 | Known for her roles as wicked matriarch and mother-in-law, she has won Filmfare Best Supporting Actress Award (1959) and Sangeet Natak Akademi Award (1961) |
| Mumtaz | 1952-1976 | Critically acclaimed highly paid actress who has bagged a Filmfare Award for Best Actress (1970) and Filmfare Lifetime Achievement Award (1996) |
| Anjali Devi | 1936-1994 | A veteran Telugu and Tamil actress well known for her mythological roles in Bollywood |
| | | continued |

Table 2 — continued

| | | |
|---|---|---|
| Sabita Devi | 1924-1996 | Supporting female actor |
| Jagdeep | 1951-2012 | Especially known for his excellent comic timing and appearances in horror movies and character roles. |
| Sanjeev Kumar | 1960-1985 | An accomplished Indian film actor remembered for his versatility and genuine portrayals of characters; has bagged National Film Award for Best Actor (1971, 1973), Filmfare Award for Best Actor (1976, 1977) |
| Johnny Whisky | 1961-1997 | Popular supporting male actor |
| Kum Kum | 1954-1973 | With her sumptuous dancing talent, she has starred with superstars of the era |
| Satyen Kappu | 1952-2007 | A remembered character actor of Bollywood films |
| Shabana Azmi | 1974-2013 | Regarded as one of the finest Indian actress of film, television and theatre proficient in a variety of genres with a record of five wins of the National Film Award for Best Actress (1975, 1983, 1984, 1985, 1999), Filmfare Best Actress award (1978, 1984, 1985), Filmfare Lifetime Achievement award (2006) and several international honours |
| Amrish Puri | 1954-2005 | Primarily remembered for essaying iconic negative roles in Bollywood and international film industries; has Filmfare Best Supporting Actor awards (1986, 1997, 1998), Sangeet Natak Akademi Award (1979) |
| Kamal Haasan | 1959-2013 | Critically acclaimed Indian film actor, screenwriter, producer, director, songwriter, playback singer and choreographer; has won a record 19 Filmfare Awards ranging across five languages, four National Film Awards, Padma Shri, one Rashtrapati Award for Best Child Artist and several other state, national and international honours. |
| Jamuna | 1954-1968 | A veteran Telugu actress who has also won Filmfare Best Supporting Actress award (1968) for a Hindi movie. |
| Birbal | 1966-2011 | A veteran comedian who has acted in 377 Bollywood films. |
| Leela Mishra | 1936-1986 | A character actress with roles varying from mothers, benign or evil aunt to comic roles; has acted in over 200 Hindi films |
| Manorama | 1941-2005 | A Bollywood character actress, acted in over 160 films, known best for her role as the comical tyrant mother or villainous roles |
| Jaya Malini | 1976-1988 | Has acted in over five different languages; known for her dance and vamp roles |
| Madhavi | 1981-1994 | Indian film actress acted in 7 languages in about 300 films |
| Raza Murad | 1965-2013 | With a rich baritone voice, he often portrays negative character roles |
| Shashi Kapoor | 1941-1999 | An award-winning Indian film actor, director and producer-Padma Bhushan |

continued

Table 2 — continued

| | | |
|---|---|---|
| Anil Kapoor | 1980-2013 | One of the most successful actors of Bollywood with National Film Award for Best Actor (2001), Feature Film (2008), Filmfare Best Actor Award (1989, 1993, 98), Filmfare Best Supporting Actor Award (1985, 2000) |
| Rajinikanth | 1975-2013 | Being one of the highest paid actors of Asia, he is a cultural icon holding a matinee idol status; has been bestowed Padma Bhushan (2000) |
| Anupam Kher | 1982-2013 | A versatile Indian actor who has appeared in nearly 450 films and 100 plays in almost all possible genres including international Oscar nominated films; honoured with Padma Shri (2004), National Film awards (1989, 2005), Filmfare awards (1984, 1988, 1989, 1990, 1991, 1992, 1993, 1995) |
| Shakti Kapoor | 1978-2012 | One of the leading villains in Bollywood movies also applauded for his comic roles; bagged Filmfare Best Comedian Award (1995) |
| Naseeruddin Shah | 1972-2013 | Considered to be one of the finest Indian stage and film actors; recipient of Padma Shri (1987), Padma Bhushan (2003), National Film awards (1979, 1984, 2006), Filmfare awards (1981, 1982, 1984, 1993, 1995, 1996, 1998, 2000, 2007, 2008), Best Actor Venice Film Festival (1984) |
| Aruna Irani | 1961-2010 | A popular supporting actress, has acted in over 300 films Filmfare Best Supporting Actress Award (1985, 1993), Filmfare Lifetime Achievement Award (2012) |
| Jairaj | 1929-1995 | A renowned film actor, director and producer; recipient of Dadasaheb Phalke Award for lifetime achievement (1980) |
| Tabu | 1980-2013 | Garnered critical appreciation for acting in artistic, low-budget films across five languages; won Padma Shri (2011), National Film Award for Best Actress (1997, 2002), Filmfare awards (1995, 1998, 2000, 2001, 2007) |
| Johny Lever | 1984-2013 | One of the most popular comedians in Hindi cinema, has won Filmfare Best Comedian Award (1998, 1999) including 13 nominations, |
| Kulbhushan Kharbanda | 1974-2013 | A popular Indian film, television actor, has been portrayed in a variety of roles ranging from a bald villian, doctor, police, hero to character roles; nominated for Filmfare Best Supporting Actor Award (1986) |
| Surekha Sikri | 1978-2006 | An Indian film, theatre and TV actress recently popular as the negative diva of telly wood, has won National Film Award for Best Supporting Actress (1988, 1995), Sangeet Natak Akademi Award (1989) |
| Anil Nagrath | 1966-2013 | Popular supporting actor |

continued

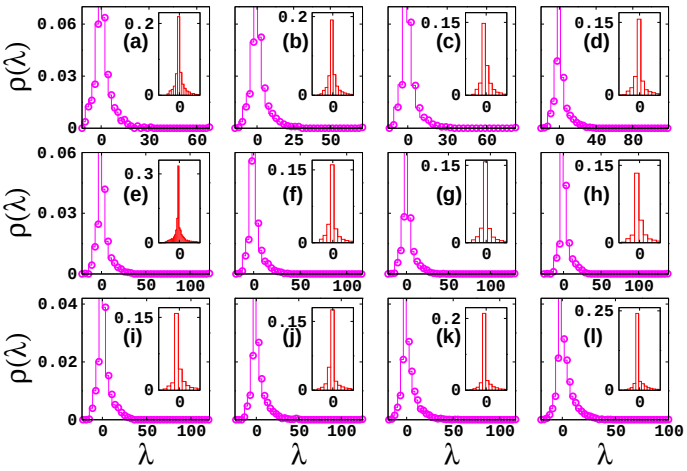
Table 2 — continued

| | | |
|---|---|---|
| Aishwarya Rai | 1997-2013 | Winner of Miss India and Miss World pageants (1994) is a leading contemporary actress of Indian cinema proficient in a range of genres; Padma Shri (2009), Filmfare Best Actress Award (1999, 2002), Most Glamorous Star of the Year (2007), Outstanding Achievement in International Cinema (2009), Decade of Global Achievement Honour (FICCI, 2011) |
| Dalip Tahil | 1974-2012 | Indian film, television and theatre actor known primarily for his negative roles has also demonstrated his versatality playing character roles in a series of national and international television serials and films |
| Irrfan Khan | 1988-2013 | India's best known international actor skilled in performing in a variety of genres; has Padma Shri (2011), Filmfare Awards (2003, 2007, 2012), Screen Actors Guild Award (2008), IRDS Film Award for social concern (2012) to his credit |
| Gulshan Grover | 1980-2013 | An Indian actor and film producer known for his villainous roles and later for his comic roles as well; has many national and international honours to his credit |
| Kashmera Shah | 1994-2011 | An Indian actress and model who has won beauty contests |
| Om Puri | 1976-2013 | Critically acclaimed for his performances in many unconventional roles in both mainstream Indian films and art films; winner of Padmashri (1990), National Film Award for Best Actor (1982, 1984), Filmfare awards (1981, 2009), Karlovy Vary International Film Festival Best Actor (1984), Brussels International Film Festival Best Actor (1998), Grand Prix Special des Amriques Montral World Film Festival for cinematographic art (1998) |
| Kalpana Pandit | 2000-2013 | An emergency physician, who turned into an Indian film actress and model; has hosted technical awards ceremony and has made red carpet appearances at Hollywood premier nights |
| Reena Kapoor | 2000-2013 | An Indian actress in films and television serials. |

## 1.3 Spectral Analyses

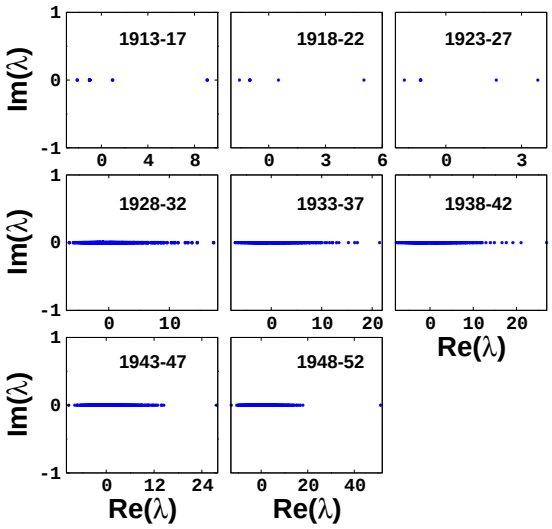
Paul Erdös and Alfred Rényi pioneered the study of random graph models [10], which persisted as a preferred method for studying networks for decades. Following this, the Barabási-Albert model of networks suggested that many complex networks follow a power law degree distribution, hence forming what is termed as scale free network, which emerged as a revolutionizing change in network analysis and completely changed the perspectives of the analysts [11]. Some of the popular networks studied henceforth namely the Internet, the World-Wide-Web, cellular networks, phone call networks, science collaboration networks etc. appeared to follow the power law distribution [8]. For the undirected networks constructed

here all the eigenvalues are real. We observe a high degeneracy at $\lambda = -1$, with almost 40% of states having this value. The presence of degeneracy at -1 is attributed to abundance of clique structure in underlying network probably arising due to several actors appearing in a same movie. Eigenvalue statistics of Bollywood network elucidate typical triangular structure, as observed for scale free networks [12, 13], with a crucial difference in having peak at -1 (SI Fig. 3).




SI Figure 3: Spectral density distribution $\rho(\lambda)$ of Bollywood networks. [(a)-(l) stand for 1953-57, 1958-62, 1963-67, 1968-72, 1973-77, 1978-82, 1983-87, 1988-92, 1993-97, 1998-02, 2003-07 and 2008-12, respectively]. Inset depicts peak of distribution.

Eigenvalue plots of Bollywood datasets (SI Fig. 4) demonstrate the presence of few eigenvalues outside the bulk region. Datasets of 1913-27 do not exhibit formation of bulk due to scarcity in number of data points. Datasets of 1928-1952 depict separation of eigenvalues from bulk indicating existence of community structure (please refer main article for elaboration).




SI Figure 4: Separation of lone eigenvalues from bulk of Bollywood datasets spanning between 1913-1952.

### 1.3.1 Nearest neighbor spacing distribution (NNSD)

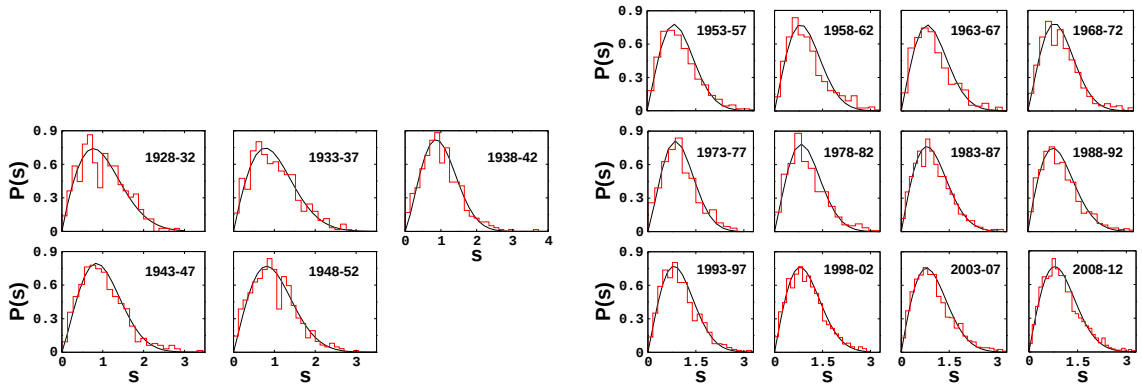
SI Fig. 5 depicts NNSD of Bollywood networks. Discussion on NNSD is provided in the main article.

### 1.3.2 $\Delta_3$ Statistics

It can be seen from SI Fig. 6 that the statistics agrees very well with the RMT prediction for some length for certain sets, and for some sets they do not follow RMT prediction of GOE statistics at all. The range for which $\Delta_3(L)$ statistics follows RMT prediction can be interpreted as providing measure of randomness in underlying network [16]. The length of the spectra which follow RMT prediction of GOE statistics is written in Table 1 of main article. In some of the sets namely 1953-57, 1958-62 and 1968-72 $\Delta_3(L)$ statistics does not follow RMT prediction at all.

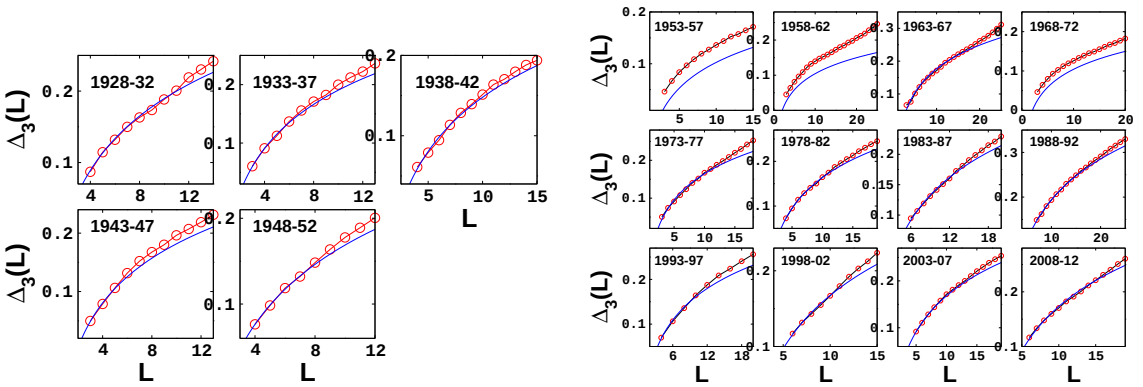
## 1.4 Net payoff

Net payoff is a measure originally borrowed from management which is modified and used as a predictive means for assessing success. PageRank algorithm has also been used to assign ranks to nodes using a Markov chain based on the structure of the graph. This algorithm was used on Hollywood datasets to uncover the relative importance of a particular actor in the graph [6]. The payoff defined here takes into account the essence of PageRank algorithm, alongwith other factors influencing the importance of a particular node. Statistics supporting the net payoff of lead male actors and female actors defined and discussed in the main article have been provided here in SI Tables 1-12. The 2003-07 span defies the trend of positive correlation between overlaps of the male actors appearing in top three consecutive positions of payoff list and their Filmfare nominations, where Amitabh Bachchan appears highest in the award nominees list. Here, it would be noteworthy to mention that the legendary Padma Shri (1984), Padma Bhushan (2001), Amitabh Bachchan (1969-2013), unlike all lead male actors of the yesteryear era, is the only one whose career never deteriorated. With 43 Filmfare nominations and being crowned as "Superstar of the Millennium" in 2000 at the Filmfare Awards, he redeems to be the superstar till date and is beyond all bounds.




SI Figure 5: Nearest-neighbor spacing distribution $P(s)$ of the adjacency matrix of Bollywood networks. Histograms are numerical results and solid lines represent the NNSD of GOE.

Lead female actors appearing in top five positions of net payoff list have been observed to bag the top three positions in terms of Filmfare award nominations (manually selected) which is very precise in the recent dataset where top five of net payoff correspond to top four nominated lead female actors, except for Katrina Kaif, who does not have any Filmfare award nomination in 2003-2007 span still appearing at the 4th position in the top five (SI Table 7). She has been one of the most popular female actors in Bollywood since 2007, net payoff seems to be predictive of her success.




SI Figure 6: $\Delta_3(L)$ statistics of Bollywood networks. The solid line represents the GOE prediction, $\Delta_3(L)$ statistics follows the RMT prediction up to length L.

SI Table 1: List of male actors holding top 10 positions in net payoff list of (a) and (b) datasets. Awards correspond to their award nominations in Filmfare in that particular span.

(a) 1953-1957

| Actors | $k$ | Award(s) |
|---|---|---|
| Ashok Kumar | 115 | - |
| Balraj Sahni | 115 | - |
| Raj Kapoor | 78 | 1 |
| Dilip Kumar | 115 | 3 |
| Shammi Kapoor | 107 | - |
| Dev Anand | 84 | - |
| Kishore Kumar | 120 | - |
| Ajit | 113 | - |
| Pradeep Kumar | 114 | - |
| Mahipal | 85 | - |

(b) 1958-1962

| List of Actors | $k$ | Award(s) |
|---|---|---|
| Ashok Kumar | 156 | - |
| Dev Anand | 115 | 3 |
| Sunil Dutt | 87 | - |
| Dharmendra | 61 | - |
| Shammi Kapoor | 114 | - |
| Manoj Kumar | 73 | - |
| Rajendra Kumar | 113 | - |
| Shashi Kapoor | 48 | - |
| Pradeep Kumar | 93 | - |
| Kishore Kumar | 97 | - |

SI Table 2: List of male actors holding top 10 positions in net payoff list of (a) and (b) datasets. Awards correspond to their award nominations in Filmfare in that particular span.

(a) 1963-67

| List of Actors | k | Award(s) |
|---|---|---|
| Dharmendra | 191 | 2 |
| Ashok Kumar | 160 | 3 |
| Manoj Kumar | 107 | - |
| Biswajeet | 115 | - |
| Shashi Kapoor | 103 | - |
| Dev Anand | 88 | 1 |
| Sunil Dutt | 110 | 2 |
| Sanjeev Kumar | 61 | - |
| Dara Singh Randhawa | 63 | - |
| Rajendra Kumar | 72 | 4 |

(b) 1968-72

| List of Actors | $k$ | Award(s) |
|---|---|---|
| Amitabh Bachchan | 178 | 1 |
| Sanjeev Kumar | 247 | 2 |
| Rajesh Khanna | 234 | 5 |
| Vinod Khanna | 179 | - |
| Shatrughan Sinha | 204 | 1 |
| Dharmendra | 221 | 1 |
| Jeetendra | 204 | - |
| Shashi Kapoor | 115 | - |
| Dara Singh Randhawa | 156 | - |
| Dev Anand | 119 | - |

SI Table 3: List of male actors holding top 10 positions in net payoff list of (a) and (b) datasets. Awards correspond to their award nominations in Filmfare in that particular span.

(a) 1973-77

| List of Actors | $k$ | Award(s) |
|---|---|---|
| Rajesh Khanna | 190 | 6 |
| Sanjeev Kumar | 234 | 5 |
| Dharmendra | 258 | 2 |
| Amitabh Bachchan | 299 | 4 |
| Shashi Kapoor | 195 | 2 |
| Shatrughan Sinha | 198 | 1 |
| Vinod Khanna | 191 | 2 |
| Ashok Kumar | 191 | 2 |
| Vinod Mehra | 178 | - |
| Jeetendra | 152 | - |

(b) 1978-82

| List of Actors | $k$ | Award(s) |
|---|---|---|
| Naseruddin Shah | 117 | 3 |
| Amitabh Bachchan | 212 | 10 |
| Dharmendra | 181 | - |
| Shashi Kapoor | 226 | - |
| Rajesh Khanna | 184 | 3 |
| Jeetendra | 195 | - |
| Raj Babbar | 179 | 1 |
| Sanjeev Kumar | 221 | 5 |
| Shatrughan Sinha | 185 | 2 |
| Om Puri | 83 | 1 |

SI Table 4: List of male actors holding top 10 positions in net payoff list of (a) and (b) datasets. Awards correspond to their award nominations in Filmfare in that particular span.

(a) 1983-87

| List of Actors | $k$ | Award(s) |
|---|---|---|
| Naseruddin Shah | 218 | 5 |
| Javed Khan | 198 | - |
| Amitabh Bachchan | 217 | 4 |
| Dharmendra | 199 | 1 |
| Anil Kapoor | 195 | 2 |
| Om Puri | 207 | 1 |
| Suresh Oberoi | 246 | 1 |
| Mithun Chakraborty | 233 | - |
| Jackie Shroff | 168 | - |
| Raj Babbar | 278 | 2 |

(b) 1988-92

| List of Actors | $k$ | Award(s) |
|---|---|---|
| Mithun Chakraborty | 302 | 1 |
| Jackie Shroff | 220 | 1 |
| Govinda | 251 | - |
| Anil Kapoor | 225 | 3 |
| Sanjay Dutt | 249 | 1 |
| Jeetendra | 196 | - |
| Rishi Kapoor | 197 | 1 |
| Dharmendra | 269 | - |
| Sunny Deol | 155 | 1 |
| Akshay Kumar | 83 | - |

SI Table 5: List of male actors holding top 10 positions in net payoff list of (a) and (b) datasets. Awards correspond to their award nominations in Filmfare in that particular span.

(a) 1993-97

| List of Actors | $k$ | Award(s) |
|---|---|---|
| Shahrukh Khan | 225 | 7 |
| Raza Murad | 296 | - |
| Jackie Shroff | 236 | 5 |
| Sanjay Dutt | 162 | 1 |
| Kiran Kumar | 324 | 1 |
| Suniel Shetty | 167 | 1 |
| Naseruddin Shah | 161 | 4 |
| Govinda | 186 | 4 |
| Mithun Chakraborty | 205 | 1 |
| Akshay Kumar | 235 | 1 |

(b) 1998-02

| List of Actors | $k$ | Award(s) |
|---|---|---|
| Shahrukh Khan | 291 | 8 |
| Jackie Shroff | 398 | 2 |
| Om Puri | 286 | 3 |
| Sanjay Dutt | 304 | 2 |
| Ajay Devgn | 249 | 3 |
| Salman Khan | 199 | 4 |
| Suniel Shetty | 246 | 3 |
| Govinda | 208 | 7 |
| Akshay Kumar | 159 | 2 |
| Mithun Chakraborty | 173 | - |

SI Table 6: List of male actors holding top 10 positions in net payoff list of 2003-07 datasets. Awards correspond to their award nominations in Filmfare in that particular span.

| List of Actors | $k$ | Award(s) |
|---|---|---|
| Salman Khan | 261 | 3 |
| Irrfan Khan | 201 | 1 |
| Jackie Shroff | 206 | 1 |
| Ajay Devgn | 228 | 6 |
| Milind Gunaji | 230 | - |
| Akshay Kumar | 326 | 4 |
| Shahrukh Khan | 246 | 9 |
| Shakti Kapoor | 315 | - |
| Kay Kay Menon | 216 | 1 |
| Sanjay Dutt | 322 | 4 |

SI Table 7: List of female actors in descending order of their net payoffs in 2003-07 span who are manually selected based on their popularity, Filmfare award nominations, income *www.filmfare.com*. Award(s) correspond to their award nominations in Filmfare in that particular span.

| Name | Net payoff | Award(s) |
|---|---|---|
| Kareena Kapoor | 0.49 | 4 |
| Priyanka Chopra | 0.46 | 4 |
| Rani Mukerji | 0.44 | 10 |
| Katrina Kaif | 0.39 | - |
| Bipasha Basu | 0.37 | 4 |

SI Table 8: List of female actors in descending order of their net payoff list of (a) and (b) datasets who are manually selected based on their popularity, Filmfare award nominations, income *www.filmfare.com*. Award(s) correspond to their award nominations in Filmfare in that particular span.

(a) 1998-02

| Name | Net pay-off | Award(s) |
|---|---|---|
| Sridevi | 0.72 | 1 |
| Rani Mukerji | 0.58 | 2 |
| Tabu | 0.54 | 7 |
| Mahima Choudhary | 0.53 | 4 |
| Aishwarya Rai | 0.51 | 4 |

(b) 1993-97

| Name | Net pay-off | Award(s) |
|---|---|---|
| Manisha Koirala | 0.51 | 5 |
| Raveena Tandon | 0.47 | 1 |
| Tabu | 0.43 | 3 |
| Juhi Chawla | 0.42 | 1 |
| Madhuri Dixit | 0.40 | 6 |

SI Table 9: List of female actors in descending order of their net payoff list of (a) and (b) datasets who are manually selected based on their popularity, Filmfare award nominations, income *www.filmfare.com*. Award(s) correspond to their award nominations in Filmfare in that particular span.

(a) 1988-92

| Name | Net pay-off | Award(s) |
|---|---|---|
| Madhuri Dixit | 0.51 | 4 |
| Juhi Chawla | 0.43 | 2 |
| Dimple Kapadia | 0.33 | 1 |
| Shilpa Shirodkar | 0.29 | - |
| Farha | 0.28 | - |

(b) 1983-87

| Name | Net pay-off | Award(s) |
|---|---|---|
| Sadhana | 0.56 | - |
| Rekha | 0.48 | 2 |
| Meenakshi Seshadri | 0.45 | - |
| Hema Malini | 0.45 | - |
| Sridevi | 0.44 | 1 |

SI Table 10: List of female actors in descending order of their net payoff list of (a) and (b) datasets who are manually selected based on their popularity, Filmfare award nominations, income *www.filmfare.com*. Award(s) correspond to their award nominations in Filmfare in that particular span.

(a) 1978-82

| Name | Net pay-off | Award(s) |
|---|---|---|
| Rekha | 0.67 | 5 |
| Sarika | 0.63 | 1 |
| Hema Malini | 0.63 | 3 |
| Parveen Babi | 0.54 | - |
| Shabana Azmi | 0.54 | 2 |

(b) 1973-77

| Name | Net pay-off | Award(s) |
|---|---|---|
| Rekha | 0.66 | - |
| Hema Malini | 0.62 | 6 |
| Reena Roy | 0.50 | 1 |
| Parveen Babi | 0.49 | - |
| Zeenat Aman | 0.46 | 1 |

SI Table 11: List of female actors in descending order of their net payoff list of (a) and (b) datasets who are manually selected based on their popularity, Filmfare award nominations, income *www.filmfare.com*. Award(s) correspond to their award nominations in Filmfare in that particular span.

(a) 1968-72

| Name | Net pay-off | Award(s) |
|---|---|---|
| Sulochana | 0.78 | - |
| Mumtaz | 0.64 | 3 |
| Hema Malini | 0.60 | - |
| Jaya Bachchan | 0.48 | 2 |
| Rekha | 0.44 | - |

(b) 1963-67

| Name | Net pay-off | Award(s) |
|---|---|---|
| Sulochana | 0.76 | - |
| Mumtaz | 0.74 | - |
| Mala Sinha | 0.48 | 3 |
| Meena Kumari | 0.38 | 6 |
| Tanuja | 0.38 | - |

SI Table 12: List of female actors in descending order of their net payoff list of (a) and (b) datasets who are manually selected based on their popularity, Filmfare award nominations, income *www.filmfare.com*. Award(s) correspond to their award nominations in Filmfare in that particular span.

(a) 1958-62

| Name | Net pay-off | Award(s) |
|---|---|---|
| Sulochana | 0.55 | - |
| Mala Sinha | 0.50 | 1 |
| Mumtaz | 0.46 | - |
| Meena Kumari | 0.40 | 1 |
| Vyjayantimala | 0.38 | 2 |

(b) 1953-57

| Name | Net pay-off | Award(s) |
|---|---|---|
| Shyama | 0.72 | - |
| Meena Kumari | 0.61 | 2 |
| Sulochana | 0.58 | - |
| Vyjayantimala | 0.51 | 1 |

# References


[1] Newman M, E, J (2001) Scientific collaboration networks: I. Network construction and fundamental results. *Phys Rev E* 64: e016131(1-8).

[2] Watts D, J, Strogatz S, H (1998) Collective dynamics of 'small-world networks. *Nature* 393: 440-442.

[3] Newman M, E, J, Strogatz S, H, Watts D, J (2001) Random graphs with arbitrary degree distributions and their applications. *Phys Rev E* 64: e026118(1-17).

[4] Neville J, Jensen D (2007) Relational Dependency Networks. *JMLR* 8: 653-692.

[5] Boldi P, Rosa M, Santini M, Vigna S (2011) Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks. *Proc of the 20th international conference on World Wide Web*, (Hyderabad, India), pp 587-596.

[6] Boldi P, Rosa M, Vigna S (2011) Robustness of Social Networks: Comparative Results Based on Distance Distributions. *Soc Inform* 6984: 8-21.

[7] Albert R, Barabási A-L (2000) Topology of Evolving Networks: Local Events and Universality. *Phys Rev Lett* 85(24): 5234-5237.

[8] Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74(1): 47-97.

[9] Newman M, E, J (2003) The structure and function of complex networks. *SIAM Rev* 45(2): 167-256.

[10] Erdős P, Rényi A (1960) On the Evolution of Random Graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5: 17-61.

[11] Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286: 509-512.

[12] Farkas I, J, Derényi I, Barabási A-L, Vicsek T (2001) Spectra of real-world graphs: Beyond the semi-circle law. *Phys Rev E* 64: e026704(1-12).

[13] Dorogovstev S, N, Goltsev A, V, Mendes J, F, F, Samukhin A, N (2003) Spectra of complex networks. *Phys Rev E* 68: e046109(1-10).

[14] Mehta M, L (1991) *Random Matrices* (Academic Press, New York), 2nd Ed, pp 2-667.

[15] Brody T, A (1973) Statistical measure for repulsion of energy-levels. *Lett Nuovo Cimento* 7(12): 482-484.

[16] Jalan S, Bandyopadhyay J, N (2009) Randomness of random networks: A random matrix analysis. *Europhys Lett* 87: e48010(1-5).